OPEN TEXT CORPORATION

Introduction

1.0 EXECUTIVE SUMMARY

Open Text Corporation is a successful high technology company in a market niche with outstanding potential for the 1990's and beyond. From a startup company a year and a half ago, Open Text has increased revenue by more than 800%, to $1.7 million in fiscal 1992 with net income of $257,000. Projections for fiscal 1993 see further increases in sales of 60 - 75% with revenues of $2.5 to $3 million.

Open Text offers technology for the management of textual data and integrated business solutions. The software comprises a text database management system with the fastest search engine available, a text display/browser, a toolkit for the conversion of text from one format to another and a Standard Generalized Markup Language (SGML) text database processor. The technology is recognized for excellence in retrieval speed, ease of integration, user-friendly interface, multi-media support, and SGML.

During 1992, Open Text successfully made the transition from a primarily academic clientele to one based on the commercial sector. A portfolio of prestigious customers now includes the Canadian Pharmaceutical Association, Union Bank of Switzerland, Mutual Life Assurance Company of Canada, National Security Agency, Grolier Publishing, Peugeot, Bibliothèque de France, and Caterpillar, to name a few. In the course of the transition to commercial sector sales, opportunities have arisen to develop associated technologies which have the potential to attain a substantial market share.

Together with the development of new technologies, has been the opening of new geographic markets. For example, with no local marketing staff, Open Text was able to negotiate significant six figure contracts with the Union Bank of Switzerland. Once the company's reputation was established, contracts were also signed with Ringier Publishing, Switzerland's largest publisher and STEP, a subsidiary of the world's leading publisher of scientific and academic journals. Similarly in the Washington area, significant contracts were signed with NSA, BTG (Jane's Fighting Ships), Pentagon, Booz Allen and others. A deliberate and concentrated marketing effort in regions where recent successes have been realized will lead to growth in market share. Specific regions have been identified for this increased marketing activity.

2.0   CORPORATE OVERVIEW

2.1   Origins - University of Waterloo

Open Text's technology was developed at the University of Waterloo in a collaborative project valued in excess of $6 million with contributions from both government sources and the private sector, notably IBM and the Oxford University Press.

In 1985, the University of Waterloo and the Oxford University Press signed an agreement to undertake a multi-million dollar project to computerize the *Oxford English Dictionary (OED)*. Given the size and complexity of the *OED* (22,000 pages, 60 million running words, 2.4 million quotations and over 80 identified structures), no existing technology was adequate for the purpose. The University of Waterloo contracted to create software that would manage the huge online textual resource (570 megabytes), allow fast, efficient full-text searching and provide flexible, attractive displays.

The technology was created and developed from the ground up, driven not only by a challenging data resource - the *OED* - but also by a standard of excellence set by the internationally respected Oxford University Press publishers and University of Waterloo computer scientists. The result of this collaboration was a family of leading-edge products, based on new technology, with the highest standards for usability. Although the original agreement to create and develop the software was with the Oxford University Press, the intention of the computer scientists was that the software would be general purpose. This goal has been realized with international installations in both business and universities for a range of applications involving information management.

Open Text's high product quality derives from the manner in which the technology was created. Each of the tools was developed under the guidance of leading research computer scientists at the University of Waterloo, dedicated to perfection and unhampered by the cost constraints and delivery deadlines typical of commercial software development. Each component of the database technology was implemented, discarded and reimplemented until the original research goals had been achieved.

2.2   Open Text Corporation 1991-1992

When it became apparent at the University of Waterloo that the software had commercial value, Open Text was formed and acquired the technology. During the first two years, the company marketed the product primarily to academic institutions and spent significant time and effort commercializing the software. By mid-1991 an impressive client list of academic institutions had been established. At that time the company entered into a new phase of its growth with the addition of marketing-focused management and the restructuring of certain underlying agreements.

The primary objectives of the newly organized corporation were twofold: to become profitable by

capitalizing on the investments made by Open Text in the commercialization of the products; and to make the transition from a primarily academic clientele to one based on the commercial sector. Open Text Corporation has successfully reached and surpassed both objectives. Open Text has been profitable each and every month since incorporation, with increases in revenue in its first year of 849%.

The company has upgraded its physical office facilities to reflect appropriate business standards with a move to the Allen Square Building in Waterloo. Management and human resources have been acquired at the highest level of professionalism. A very competent group of experienced professionals has been attracted to Open Text and melded into a small but extremely hardworking team which will form a stable nucleus of an expanded company in the months and years ahead. Open Text grew from four to eighteen employees by the end of 1992.

Open Text continues its close relationship with the Centre for Text Research at the University of Waterloo by providing grants to fund core research into maintaining the Open Text products on the cutting edge of technology.

The marketing activities of the company can be summarized by noting that 1991- 1992 was a year of transition. From being a minor player in the academic market with an unidentified technology, Open Text has established a presence with high profile customers in key vertical markets and a growing market understanding of our technological benefits.

Professional sales/marketing techniques were established with new sales aids, demonstration equipment, sales booth and detailed sales strategies. Open Text participated with Delphi at industry trade shows to establish industry credibility in Los Angeles, Boston, Dallas, Chicago and Ottawa, among others. Many other trade shows brought attention to our product line and its unique capabilities.

Reference accounts were obtained in several key industry areas including:

Financial Market: Union Bank of Switzerland, Mutual Life of Canada
Government: Ontario Legislature, National Security Agency
Pharmaceutical: Canadian Pharmaceutical Association
Automotive: Peugeot, Caterpillar
Publishing: Thomson, Ringier, Grolier
Aerospace: Pratt & Whitney

The operations and development activities have similarly advanced. Customer support was redefined and a new system implemented. A Business Partner Handbook was developed and formal training programs instituted. A Product Scheduling System was put in place. Documentation needs were defined and documentation has been produced. An interim release was sent to customers and PAT 4.0 completed and released. Among other development programs, PC PAT was brought to the prototype stage as well as a CD-ROM version which has unique capabilities.

2.2.1   Client Summaries

The following overview of client applications illustrates Open Text's successful transition in 1992.

Union Bank of Switzerland

Union Bank of Switzerland utilizes Open Text technology to optimize information flow to senior executives concerning their multi-national clientele. Information is accessed from a variety of financial newspapers, the Dow Jones newswire and internal documents. The project is successfully meeting its goal of being used effectively by senior executives after a single fifteen-minute training session.

Ontario Legislative Assembly

The Ontario Legislative Assembly maintains and accesses the daily legislative proceedings and its *Precedents* with Open Text's text database management system. Fast and efficient retrieval of information is now available to hundreds of users accessing the database through a wide range of front-ends.

University of Michigan

The University of Michigan has used Open Text technology for the past three years to provide library reference works to an entire academic community. The system offers access to a wide range of materials over campus-wide networks.

The Mutual Group

Mutual Life of Canada has based the automation of its translation department on Open Text's products. The Open Text system achieved a full return on investment in less than twelve months.

Peugeot Incorporated

Peugeot Incorporated has implemented an advanced automobile repair system built around Open Text's technology. The system incorporates text, image, tabular information and diagnostics based on SGML.

Ringier Publishing

Ringier Publishing maintains a picture file of over seven million images. Locating the appropriate picture, identified by title, caption, or any other field, is a simple and fast process with Open Text's search and retrieval technology.

Grolier Publishing

Grolier Publishing is utilizing Open Text's software as part of an integrated editorial system to retrieve information from a master datafile. This system includes relational database technology, sophisticated colour graphics and workflow control.

Japan Electronic Dictionary

Japan Electronic Dictionary, a $100 million Government of Japan project, has purchased Open Text technology to manage the publication of reference works. The project will result in the provision of support for Oriental character sets which in turn will lead to many opportunities in the Japanese market.

2.3 The Future: 1993 and Beyond

The market for text retrieval products exceeds $225,000,000 and is growing at a rate of 30% per year. To address this market in future years, Open Text Corporation is directing its activities to leverage its sales resources to the maximum extent. It will achieve this by the following methods. First, Open Text will enter into an increasing number of Business Partner relationships. This is essential to a "reseller" arrangement whereby a local system integration company sells and supports the Open Text Product. Several successful Business Partner relationships have already been concluded with such companies as Booz Allen & Hamilton, SRA, Xyvision (end user Grolier Publishing), B.T.G. (end user Jane's Fighting Ships), A.I.S. (end users Peugeot, Bibliotheque du France), Public Sector Systems (end user DSS - Ottawa), Computer Science Corporation (end user Caterpillar) and Genesis (end user National Security Agency).

The second method of leveraging sales activity is to focus an effort on concluding a number of Strategic Alliances with corporations who will embed the Open Text technology into their product and thus Open Text will benefit from a royalty payment every time the product is sold.

The third strategy is to expand our markets internationally outside our current areas of activity. The European and Japanese markets appear to be very promising and by having reference accounts already in both areas we are well positioned to seek out and enter into business relationships with local based national companies.

The existing technology will evolve in different ways over the next few years. Open Text has developed and implemented a strategy to ensure that its products remain on the leading edge of technology.

The existing products of Open Text will be integrated with allied and complementary technologies to offer a broader solution to the end user. In this regard Open Text has approached several other leading Canadian corporations to determine their interest in pursuing the development of an integrated product.

The second area of evolution of the Open Text technology involves developing new technology as well as enhancing the existing technology so that our full text retrieval tools can interface directly with most existing relational database management systems. This market is several billion dollars in size and such a system would represent a major breakthrough in the industry. To achieve this objective within a two to three year time frame, Open Text has contracted with the University of Waterloo for research and development. The project is headed by Dr. Frank Tompa, the Chairman of the Computer Science department at the University of Waterloo.

Concurrently the company has submitted a $9 million project proposal to the Canadian Government to obtain funding for such a project. Grafnetix Systems, Public Sector Systems, Fulcrum Technologies, Megalith Technologies, Softquad, Exoterica, InContext and the University of Waterloo have committed to participate and invest in the project and The Strategic Technology Fund approved an initial $50,000 grant to assist with the costs of defining the combined technologies.

Utilizing the methods above Open Text has ensured that its technology will be broadened, its product line substantially enhanced, and it is investing in a very cost effective manner to ensure that its technological future leads the market.

3.0 TECHNOLOGY

3.1 Text DBMS

.... increased communication and search speed are the two most important criteria when implementing a text retrieval system ... survey results, 1992 Delphi Report, p.39

.... client server architecture tops list of 24 key functionalities for text retrieval systems ... survey results, 1992 Delphi Report, p.42

The innovative text search and retrieval schemes designed and implemented by Open Text Corporation are oriented toward high speed searches of large databases and search and retrieval of structured text, not simple document retrieval. Open Text Corporation's text DBMS is based on string search technology, not traditional inverted wordlist approaches. This string search technology has been designed specifically for a text database running on today's systems, with today's memory, CPU, disk and network tradeoffs.

3.1.1 Characteristics of String Search

- Search performance is independent of the query size. Regardless of whether one word or many sentences are being searched for, database response time shows no degradation.

- Response time is constant regardless of the number of hits in the database. Whether there are millions of hits, or just one, the Open Text text DBMS retains its rapid response.

- Search speed is relatively insensitive to the size of the database:

Table 1:

Size of Database	Response Time for String Search (Seconds)
100 Mbytes	1 second
1 Gbyte	1.5 seconds
10 + Gbytes	2.0 seconds

Open Text's string search technology does not require common words to be ignored ("stopped out") during the search process. Although it is rare that a user would want to search for common words such as "the", "and", "of", and "about" by themselves, systems that ignore these words suffer a dramatic loss of context information. For example, using string search technology, a search on the phrase "the association of" may result in two or three hits in a given database. Using inverted word technology, however, the words "the" and "of" would have to be ignored. This would result in finding all occurrences of the word "association", resulting in thousands of database hits for this query. It is evident that valuablecontextual information has been lost by ignoring these so-called common words.

Native structure handling support by Open Text's DBMS allows text access below the document level. Such component based search and retrieval is essential for a true database approach to text.

The capabilities described allow Open Text's DBMS to support both Standard Generalized Markup Language (SGML) and non-SGML texts.

Since many users are currently in the planning phases for SGML, Open Text's concurrent support for both SGML and non-SGML data helps ease the migration path to SGML.

3.1.2   Open Text's Text DBMS Architecture

.... the top functional categories for a text retrieval system are architecture and   integration ... survey results, 1992 Delphi Report, p. 40

Traditional players in the text retrieval industry are based on:

* old inverted word list search technology
* monolithic designs current at the time of their implementation

Open Text's text DBMS has been designed for:

* deployment in multi-vendor solutions
* client/server architectures
* ease of integration

- replaceable modules

The Open Text DBMS exposes and documents what are traditionally   internal   system communication paths.

3.2 Products

Open Text's suite of text management tools include: the PAT database system, the LECTOR text display system, TTK, the Transduction Tool Kit, and the SGML Text Database Processor. These tools have been designed together and are easily integrated in a variety of ways. Moreover, they are designed for easy integration, separately or together, with other software packages, facilitating the provision of solutions to a wide range of business needs.

The tools have been tested and refined and are in active use at customer sites.

3.2.1 The PAT Text Database System

PAT is a text management system that enables fast, efficient, multi-user access to and rapid update of, large, text-dominated databases.

The following list outlines the features of PAT which make it unique in the marketplace:

- The fastest text searcher available, capable of searching through hundreds of millions of words per second.

- Searches for words, parts of words, or long phrases with equal speed, whether there are millions of matches or only a handful, making possible a wide variety of queries which other systems cannot handle.

- Effectively handles the natural structure of text including the irregular, nested hierarchies commonly created by human authors.

- Requires no special input format, easily adapting, for example, to a variety of common proprietary editing packages.

- Applied to any file, in any format.

- Unmatched in the ease with which it can be integrated with other software components.

- Designed to any open systems client-server architecture, allowing it to operate as a distributed database.

- Allows parallel searching of multiple databases located in a single site or around the world.

- Flexible, intuitive, easy-to-use OSF/Motif and Microsoft Windows user interfaces that have independently been rated as the best in the industry.

In addition to the features of PAT, it is worthwhile to review some of its applications:

- Editorial control of large reference works such as dictionaries andencyclopedias.

- Research in linguistics and on the structure of language.

- Provision of enterprise-wide access to general information resources.

- Administrative tracking of large volumes of electronic correspondence, originating on a broad range of computers located around the world.

3.2.2   The LECTOR Text Display System

LECTOR allows rapid real-time display of text, formatted to match the original or target document. It was originally designed to display the results of PAT database searches, but has been used as a front end for many different text applications where fast, high-quality information display is required.

LECTOR increases productivity by applying standard typographical techniques such as multiple columns, typefaces, spacing, outlining and colour. It employs a set of user-written, customizable style sheets to control the format of the display text. It is possible to generate multiple style sheets which allow parallel but different views of the same text. For example, LECTOR can be used to switch between a standard format, an outline format that suppresses all text except for titles and subtitles and an index format that suppresses everything but index items. The outline format can be used to quickly relocate within a document or scan its entire contents. Facilities for paging, scrolling and repositioning enhance this capability.

LECTOR uses either OSF/Motif or Microsoft Windows and can turn any workstation into a state-of-the-art information delivery system.

3.2.3   SGML Text Database Processor

Open Text/SGML is based on Open Text's advanced PAT text database technology, and provides all the advantages of the PAT system:

- High performance, independent of the number of matches to a search

- Ease of integration with other software through an advanced high- level Application Programming Interface

- International application, using the Open Text/WorldSearch product

- Large, distributed database access, using the Open Text/ATLAS product

- A suite of highly-praised Graphical User Interfaces for Microsoft Windows and OSF/Motif, including interfaces for both power and casual database users

Open Text/SGML is based on Open Text's PAT text database manager. PAT is well-suited for this task, since its internal structure-handling facilities are exactly those required for SGML processing.

To automate the processing of SGML, Open Text/SGML includes the following specialized modules:

- A preprocessor which applies an SGML parser to the instance, including DTD, and automatically generates all the required structure indices.

- A "Direct SGML Query" Graphic User Interface facility, that supports queries phrased directly in terms of elements, attributes, and attribute values.

- A tagged-text formatter (Lector), based on user-written style sheets, that formats SGML-tagged text onto the screen, using font changes, colour, text suppression, and white space, at a speed which makes it suitable for use as a database browser.

The value of SGML has been recognized, and a lot of enterprises are dedicating very significant effort to moving towards it. Open Text/SGML, for the first time, provides a high-performance, easy-to-use repository for storing and retrieving SGML text without compromising either the information's structure or the user's productivity.

3.3 Technological Advantages

The Open Text technology comprises a suite of tools capable of handling all aspects of text database management. Each component is at the leading edge of technological development, offering features that are not available in other products. The tools work together or separately , are easily integrated with other software solutions, and can be adapted to a variety of client-server architectures.

Every objective comparison between Open Text tools and other technology in the field has revealed Open Text tools to have a substantial technical lead in speed, power, structure-handling, flexibility and user interface. These were developed by a close-knit team of world-renowned scientists and engineers who shared common goals to address different aspects of the text management problem.

The following describe advanced technological options that are available with Open Text's software.

Large Databases

Open Text's products are particularly well suited to text search and retrieval on large text databases. This ability to handle large databases stems from the algorithms fundamental to the software design. The key point is that search response time is logarithmic in the size of the database. Consequently, response time for databases larger than 100 MBytes is near constant. This constant figure is in the range of .75 to 1.5 seconds, depending on the computer hardware. This insensitivity to database size can be extremely critical for databases which range in the neighbourhood of hundreds of megabytes to gigabytes in size.

Graphics Handling

The handling of graphics is also readily accomplished due to the modular nature of Open Text's product design. By exposing what are in many systems considered to be "internal" communication paths, Open Text allows users and system integrators to insert their own graphics or filters. These filters can then be triggered by selected query results to display arbitrary graphical data.

Integration with SGML Editors

Open Text's text search and display software has been closely integrated with SGML editors on numerous occasions. Applications range from publishing houses such as Grolier Encyclopedia to internal editorial applications such as Thomson Publishing. In these environments users can perform text searches on SGML data rapidly, using Open Text's powerful structure handling. Once the desired SGML components have been retrieved, they can be edited using the SGML editor. Alternatively, Open Text's LECTOR product can be used for rapid viewing of the documents. Integrated environments such as these have been readily created in conjunction with products such as SoftQuad and Arbor Text'sSGML editing environments.

Adherence to Standards

Main considerations in the development of the tools were portability, reliability and adherence to industry standards. Each of the tools has been coded in the "C" language and will execute in any environment that conforms to a Unix standard, including System V, X/Open and Posix. Networking and client-server communications are also based on the industry standard TCP/IP networking protocol allowing operations over popular LANs (local area networks) such as Ethernet, Token Ring, Novell and Banyan, as well as ISO-standard WANs (wide area networks). The graphical user interface modules adhere to industry standards including the OSF/Motif and Microsoft Windows. Finally, while the tools operate on any data file, they are ideally suited to manage databases that conform to the non-proprietary ISO-defined SGML (Standard Generalized Mark- up Language) which was explicitly designed to enable the open interchange of information.

Client-Server Architecture

Open Text's tools are designed to operate in the client-server model. Each of the modules - database management, data conversion and user interface - executes independently of its own process, and exchanges commands and data with other modules using text-only messages on standard interprocess communication channels. The format and content of the messages is completely documented for purchasers of the tools. Since the modules execute independently of

one another, regardless of their role as server or client, the Open Text tools are easily connected with user-written custom tools and products from other vendors, facilitating the creation of distributed multimedia applications.

Open System Orientation

Open System architecture allows maximum flexibility in the use of hardware and software to provide appropriate and innovative solutions. A system is considered open when it can be easily combined with other tools to create a solution, applied to data to accomplish tasks, and run on an existing setup without the purchase of special equipment. With its compliance to industry standards, client-server construction, SGML support and ability to adapt to a variety of input formats, Open Text technology provides an extremely open product set, achieving the goal of excellence established by its original creators.

3.4 Evolving Product Potential

During the past year, Open Text has made a successful transition from a primarily academic clientele to one based on the commercial sector. Opportunities have arisen during the transition to develop new products based onthe core technology.

3.4.1 Parallel PAT

Parallel PAT's development was motivated by the need to access very large databases. Parallel PAT has proven to provide an extremely fast and consistent level of performance on very large databases that is simply not attainable using any other currently available technology. Further market feedback suggests that it has another application area that is at least as important. This is among those organizations which are large enough to have several workgroups which are each building and maintaining their own information pools.

On the one hand, such groups may be widely separated organizationally and have no mandate to co-operate to service enterprise database needs. On the other hand, it may be advantageous to the organization to have parallel access to more than one such database at a time; potentially, to all of them. Parallel PAT meets the needs of this scenario perfectly. None of the groups need make any effort to co-exist or co-operate with any of the others, but a corporate virtual database may be defined which performs global parallel search on any permuted subset of these database instances.

3.4.2 Integrated T/RDBMS

Open Text is lead participant in the nine-member Canadian Strategic Software Consortium's $9 million project to create integrated T/RDBMS technology. The consortium was founded last September when Industry, Science and Technology Canada approved a $50,000 feasibility study for which Open Text Corporation was the lead applicant. Members

of the consortium now include the University of Waterloo, Fulcrum Technologies, Megalith Technologies, Softquad, Exoterica, InContext, Public Sector Systems, and Grafnetix Systems Inc.

The main goal of the consortium is to provide integration of text and relational database technology as well as related technologies, with the development of candidate standards. Integration of database technologies will take place at a lower level, while other technologies will be integrated at a higher level, perhaps through the user interface.

Core research will be carried out at the University of Waterloo by world leaders in database research, Professor Paul Larson, Professor Frank Tompa and Professor Gaston Gonnet. Research and Development will be conducted concurrently by the private sector members of the consortium.